Annual ADFSL Conference on Digital Forensics, Security and Law

2007
Proceedings

The Gap between Theory and Practice in Digital Forensics

Joseph C. Sremack
LECG, Washington, DC USA

The Gap between Theory and Practice in Digital Forensics

Joseph C. Sremack
LECG
Washington, DC USA

ABSTRACT

Digital forensics is a young field that is being defined by the reactive nature of its development – in terms of both research and practice. As technology develops, digital forensics is forced to react and adapt. The rapid development of technology and the lack of an established theoretical foundation has led to a disconnect between the theory and practice of digital forensics. While the base theoretical issues are being worked on by researchers, practitioners are dealing with entirely new sets of issues. The complexity of investigations is increasing, and anti-forensics techniques are advancing as well. The disconnect will be resolved by economic and legal factors, as well as each side understanding their role in the development of this field and improving their channels of communication. This understanding will lead to digital forensics becoming a more mature and effective field.

Keywords: digital forensics, theory, research, practice

1. INTRODUCTION

Digital forensics is a rapidly developing field that continues to evolve, both in practice and in research. Technology continues to advance, as does the sophistication of anti-forensics techniques, which forces digital forensics researchers and practitioners to adapt. Moreover, rules for admissibility are becoming more stringent because the legal community is continuing to better understand technology and digital forensics. The legal community is requiring that digital forensics become more scientifically rigorous, much like traditional forensics.

Digital forensics is still in its infancy, especially when compared to other areas of forensic science. The term "computer forensics" was first defined in 1991 by the International Association of Computer Investigative Specialists (IACIS) [8], and the term has since evolved into "digital forensics" to encompass all digital media. Various U.S. government agencies have performed digital forensics since the mid-1980s; the FBI's Computer Analysis and Response Team (CART), for instance, was formed in 1984 [6]. The field is, therefore, approximately 20 years old. Compare this to traditional forensic science, where the field, laboratory, and medical disciplines have a rich history that extends back to the 19th century.

In its beginning, the need for digital forensics was immediate, and the development of digital forensics techniques preceded any academic research or theoretical backing. Digital forensics practitioners developed best practices that were court-admissible on an ad hoc basis. The practitioners had no choice but to assume a reactive stance and perform investigations to the best of their abilities, without a theoretical framework or established legal principles. Digital forensics researchers had not yet stepped in to provide their input. This reactive stance and the lack of a theoretical framework meant that digital forensics evolved based on the specific needs of investigations and not necessarily on theoretical soundness.

Digital forensics research was initiated in the mid-1990s and evolved into a community of researchers with peer-reviewed journals and conferences. The research was born out of U.S. federal agencies' needs to standardize and formalize the digital forensics process. One such conference was the International Law Enforcement Conference on Computer Evidence, first held in 1993 [10]. These conferences focused on the problems facing investigators, including how to handle non-computer-based digital evidence and how to standardize the investigation process. The research community has continued to evolve with the introduction of conferences and digital forensics-specific journals like the

Digital Forensics Research Workshop (DFRWS) and the International Journal of Digital Evidence (IJDE). These journals and conferences recognized the need for peer-reviewed digital forensics research. The academic scrutiny of current techniques and methodologies – as well as new ones – is critical for the advancement of the field.

Through their extensive research in digital forensics, researchers have developed numerous theoretical approaches to this field, but a growing disconnect between practitioners and researchers is occurring. The research has produced advancements in steganography, file system analysis, data reduction, and other areas. These advancements have improved digital forensics, but the difficulty is that some of the research is not in line with the needs of practitioners. This is caused by the fact that most practitioners are unaware of the research community, and many researchers are not fully aware of the technologies and techniques employed by practitioners.

This paper explores the various gaps between theory and practice in digital forensics. It also describes several forces that will "plug the gaps" and improve the overall field of digital forensics. In the first section, the reactive nature of digital forensics is described. The second section lists the goals for both theory and practice, explains how those goals are orthogonal to each other, and how that affects digital forensics research. The third section looks into some of the most important current issues faced by practitioners and the progress made by researchers. The fourth section examines the gaps between theory and practice, and the final section examines the various drivers that will lead to improvement in digital forensics.

2. REACTIVE NATURE OF DIGITAL FORENSICS

As a field, digital forensics reacts in various ways. When an incident or investigation occurs, digital forensics is performed. When new technology is introduced, digital forensics adapts. When case law changes, so too does digital forensics. Digital forensics is thus highly reactive.

Practitioners have been in a reactive posture since the inception of digital forensics, and that stance has been the basis for the development of digital forensics best practices and techniques. Digital investigations in the 1990s were typically conducted with primitive forensics software, few resources, and scant literature or best practices. Practitioners have developed best practices of digital forensics by solving immediate problems in the best, most time-permitting manner possible. These best practice techniques and methodologies have over time formed the literature of digital forensics. That is, the reactive techniques have evolved into the best practices and theory for lack of academic research.

The reactive nature of digital forensics has made the formation of a digital forensics science difficult. Several researchers have noted this problem. Gary Palmer claims that computer forensic analysis is not a true science, since the established methodologies and techniques are based on reactions to practical needs rather than sound scientific principles [9]. Adding to this confusion are several misconceptions. Traditional forensic science is founded on sound scientific principles from the soft science of biology and the hard science of physics. The principles and methodologies of traditional forensic science have been thoroughly scrutinized over the past hundred years in both academic and legal circles. During this time, both sides realized that no evidence is irrefutable, and that investigations' conclusions must be convincing when faced with counterexamples or questions as to the soundness of the analysis. The scrutiny has led to a well-accepted body of literature for traditional forensics. Digital forensics, however, is a relatively new field. Computer crimes began to occur before technical and legal researchers truly understood the underlying problems of analysis methods and the validity of such analysis. This lack of understanding has resulted in theory that is founded on a series of techniques and methodologies that are designed to analyze specific types of cases.

Reactive problems run counter to scientific research. Research is better suited for the solving of problems where there is a positive, proactive solution, rather than mitigatory and recovery-based ones. Reactive problems are typically thought of in terms of purely practical or engineering-based. This notion is because most reactive problems involve configuration-specific issues. While digital forensics

research is much more than handling specific system configurations, the literature has largely evolved in that way. For example, there are Windows-specific methodologies and Unix-specific methodologies. Both of these methodologies grew from specific needs to find and analyze data that reside in different formats. The literature did not evolve from a generic, unified one.

Another difficulty with the reactive nature of digital forensics is that practitioners, at least partially, ignore the research community. Practitioners are typically under strict time constraints for completing their investigations. They care about successfully completing an investigation and not about mathematical formalizations or other theoretical issues. Most research does not apply to their investigations, and the research that does apply is typically not practically described or presented.

3. GOALS OF RESEARCHERS AND PRACTITIONERS

In most fields, researchers solve problems that face practitioners, and practitioners rely on researchers for solutions to their problems. If practitioners do not rely on researchers in some way, then researchers do not serve much of a purpose. It is therefore imperative that researchers understand the needs and goals of practitioners so that the right problems are solved. It is this issue that is most critical for digital forensics researchers, for perhaps the largest gap between digital forensics researchers and practitioners is that of the practitioners' needs and the researchers' goals.

The roles of practitioner and researcher are important for both sides to understand. Research is not solely confined to academia, and conversely, practice is not confined to law enforcement and the private sector. For purposes of this paper, a practitioner is anyone who actively performs digital forensics in order to participate in a criminal or civil investigation, or to otherwise respond to an incident. A researcher is anyone who someone who innovates new tools or techniques, or refines existing tools or techniques. Figure 1 highlights the three digital forensics domains and denotes the fact that they interact and overlap.[1]

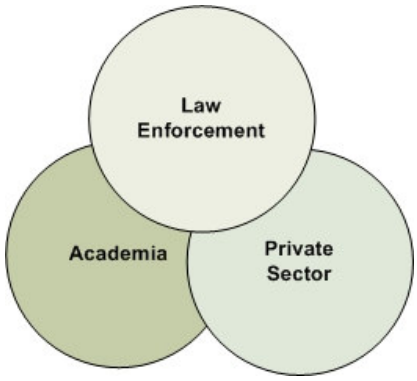

Figure 1: The three groups in digital forensics.

In conducting their work, the aim of practitioners is to perform complete, accurate, and timely investigations that are court-admissible. Each investigation must be completed in a timely manner in order to preserve data and meet court-imposed deadlines. Large volumes of data slow down any practitioner, so the data set must be reduced to eliminate non-useful data. Practitioners continually face the data reduction problem. That is, reducing the volume of data down to a manageable and yet meaningful amount. The large data sets can overwhelm investigators and can make completing investigations difficult, if not impossible. This problem is primarily responsible for practitioners relying on automated tools to cull the data set. Despite its importance, this issue is rarely discussed in digital forensics research literature.

[1] This paper does not address the distinctions and differences between researchers and practitioners from each of the three groups.

The court-admissibility of evidence and findings is critical for practitioners. Every step a forensics investigator makes can affect the admissibility of evidence. Chain of custody, data integrity verification, and the data acquisition method can all be called into question by opposing counsel, thereby potentially leading to critical evidence being deemed inadmissible. Practitioners will therefore only employ techniques that are accepted by the courts. This is based on case precedents, so any untested techniques will not be used unless necessary.

The completeness and accuracy of an investigation is also important. Every investigator aims to discover as much useful evidence as possible, corroborate that evidence, and then present the findings in the most convincing manner possible. These goals center on an investigator's ability to sift through massive amounts of data, discover the important pieces of evidence, and rule out meaningless data. The practitioner achieves these three goals by utilizing software tools to automate the process, or at least provide semi-automated analysis. An additional factor is non-data evidence, such as custodian and suspect interviews. These investigatory aspects tend to fall under the purview of investigators moreso than researchers, since researchers often ignore the issue of linking non-data evidence to data evidence.

These three goals are the reasons for why practitioners rely so heavily on commercial software. The commercial software products are known to work and have been tested in court in prior cases. If the software is called into question, counsel can cite prior usage in its defense. The commercial software packages are also designed specifically to solve digital forensics problems. They store chain of custody information, they produce meaningful reports, and they can assist with data reduction. The argument can also be made for open source software packages [1]. These packages have also been tested in court and assist with data reduction and court-admissibility. The general trend, however, is that practitioners will opt for an all-in-one Windows-based software product over a more low-level product or technique.

The goals of digital forensics researchers have been stated several times in the literature. The attendees of the Digital Forensics Research Workshop in 2001 set out the following goals and guidelines [4]:

> "The majority of current computer forensic analysis is focused on assisting the law enforcement community. The criteria that define suitability for forensic evidence in this area are the most clearly defined since computer forensic analysis must follow the same longstanding statutory and regulatory guidelines imposed on other, more traditional forensic disciplines. Existing technologies and those that are evolving, in support of law enforcement, will come under increasing scrutiny as technical knowledge expands in scope. For this reason, it is imperative that sound research steeped in the scientific method becomes fundamental to the discovery and enhancement of all tools and technologies employed to assist the courts, including digital forensic evidence."

The group goes on to say:

> "[T]o be effective, fundamental digital forensic research must provide suitable solutions with the widest possible applicability to Homeland Security. To do that the focus must be the foundation science at the root of the technologies we aim to analyze."

In other words, digital forensics research must follow existing legal guidelines and must be performed according to the scientific method. The current technologies should be evaluated according to these principles, with an eye on developing technologies. The end result should be the discovery, evaluation, and enhancement of practical, scientific tools and techniques.

The goals of researchers and practitioners are largely orthogonal. Whereas practitioners work to solve immediate problems under tight time and resource constraints, researchers work to solve deeper problems without the time and resource constraints. The types of problems they work to solve are distinctly different. Researchers are afforded the luxury of time, which allows them to more formally

understand data, such as where they are stored and for how long.

Researchers tend to focus on full knowledge during an investigation rather than the pragmatic concerns of practitioners, such as time restrictions and very large data sets. Digital forensics researchers are typically computer scientists, so most digital forensics research is that of a computer science nature. For example, many of the research papers focus on computer science issues such as formalizing investigations and using mathematical formulas to describe digital forensics. Researchers prefer to formally describe and prove an investigation, rather than discussing pragmatic problems and legal constraints.

The fact that researchers work in a different environment than practitioners should in no way limit the effectiveness of their research. Practitioners can learn about the nature of data from such research. Researchers are afforded more time, which allows them to describe best practices. This can provide practitioners a basis for their analysis, as opposed to an ad hoc or purely practical approach. Moreover, practitioners must understand digital forensics at a deep theoretical level if they wish to be experts and be able to defend their investigations in court. Any digital forensics investigator who takes the stand in court must be able to withstand cross-examination and defend her findings. The research community can be extremely valuable in this respect by defining the error rates, steganography theory, etc.

The most valuable benefit of researchers is their ability to construct scientific principles, which is critical for court-admissible evidence. The basis for the admissibility for scientific evidence in the U.S. is known as the Daubert Test for admissibility. The Daubert Test originates from *Daubert v. Merrel Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993)*, whereby evidence is only deemed admissible if it is relevant and rests on a reliable foundation [2]. The criteria for Daubert are the following:

- Has the theory or technique been reliably tested?
- Has the theory or technique been subject to peer review and publication?
- What is the known or potential rate of error with the method used?
- Has the theory or technique been generally accepted by the scientific community?

Researchers can provide these tests, metrics, and peer reviews in order to satisfy the Daubert requirements. Currently, such error rates and testing are largely absent in digital forensics research literature.

4. CURRENT ISSUES

Practitioners and researchers each have their own understanding as to what is important for digital forensics. Practitioners focus on immediate issues that impede their ability to perform quick, accurate, and complete investigations. Researchers, on the other hand, focus on formalizations, standardization, and other theoretical issues. This section outlines some of the key issues for practitioners and researchers and explains how each perceives the other's concerns.

The data reduction problem is the largest issue facing practitioners. Data storage is getting bigger and cheaper, and as such, the volume of data in digital forensics investigations is growing rapidly. Investigators have to adapt with new tools, techniques, and more computing hardware. Data mining researchers have worked for many years to develop data reduction techniques and tools, but most of this knowledge has not yet been applied to digital forensics research. Practitioners would be much more efficient if effective data mining tools and techniques were available.

Another big issue for practitioners is that of network data. It is almost unheard of to have a digital device that is not in some way networked. Like all other forensics evidence, network data must be acquired in a forensically sound way and then properly analyzed. The difficulty is collecting the volatile data and proving that the data was not altered. This is the key issue in network forensics, which researchers have begun to explore. This area is of critical importance to practitioners.

A third issue is the fast analysis of files for hidden data, i.e. steganography. When steganography is combined with the large volumes of data in most investigations and tight time deadlines, investigators simply cannot uncover all hidden data. Researchers have done much work in this area, which includes both individual file steganography and root kits. The caveat is that researchers focus on completeness instead of timeliness.

Table 1 displays several of the main issues faced by practitioners and the relative activity to solve that issue by researchers.

Practitioners' Issues
1. Blind Steganography
2. Network forensics
3. Rootkit detection
4. Data Reduction Problem
6. Handheld devices
7. Error rates for tools

Table 1: Several of practitioners' main issues

Researchers have focused more on the foundational issues in digital forensics. The main issues being addressed are formalizing the analysis process and performing automation. The goal is to create a common language for computer forensics and also a standardized methodology for performing investigations. This goal is founded on the need for consistent, reproducible analysis.

Table 2 displays several of the main issues on which researchers are focusing.

Researchers' Issues
1. Formal Methodologies
2. Root Cause Analysis
3. Automated Analysis
4. Formal Notation for Evidence

Table 2: Several of researchers' main issues

5. THE GAP BETWEEN THEORY AND PRACTICE

The main gap between the practitioners and researchers is a lack of communication. Researchers are not fully aware of the types of problems facing practitioners or the full scope of technologies they use. Practitioners, in turn, are not aware of the work done by researchers and their technical insights. Without understanding the other half, neither practitioners nor researchers can benefit from the other. This limits the field of digital forensics. There needs to be a better understanding of each side's goals, challenges, and current work.

The redundant solving of problems is one particular area that points to a lack of communication. Researchers and practitioners are both good at spotting problems. The main difference is that researchers tend to be more thorough in how they solve the problem, whereas practitioners tend to solve problems faster. Another difference is that researchers focus more on the creation of a solution

rather than its full development.

Another area of confusion is the practicality of digital forensics research, such as research for formalizing digital forensics. Much of the current work centers on the formalization of methodologies. This includes employing mathematical notation to describe a forensic investigation. This makes sense within the realm of computer science, since many digital systems can be formally modeled. The formalized models break down, however, when human actions are incorporated. What use do mathematical models really serve a digital forensics practitioner? Will he go through the same mathematical rigor to validate and verify an investigation that a NASA programmer does for space shuttle software? Hopefully not. Most mathematical models are demonstrated in a paper with an example that has little resemblance to the complexity of a real-world investigation. The complexity of most real-world investigations is fairly great. Introducing formal verification models adds to the workload and will most likely yield little benefit in terms of finding mistakes or undiscovered causal links. Also, most digital forensics practitioners are not trained in set theory. More importantly, most juries, judges, and lawyers are not trained in set theory, which means that the formal model will do more to confuse than to convince.

The goal of research in any field is not always practicality. Researchers always strive to improve existing work or tread new ground. That aim is not always in line with what "industry" does. The medical industry is a prime example of this phenomenon. Researchers often work on curing a disease or condition with treatments that may not be ready for the public for over ten years. This is not practical in that the treatment cannot immediately go to market and may not even be effective in treating the disease in the end. Still, the researchers stay in line with the long-term goals of the industry and follow the established research methodology. Not all digital forensics research stays in line with an established methodology or the long-term goals of practitioners.

Digital forensics research could be made more practical through understanding what digital forensics aims to achieve. Digital forensics practitioners need a theoretical foundation for the field. Courts demand scientific soundness and credibility for any evidence entered into court. The soundness and credibility are evaluated based on Daubert. Researchers can establish this scientific foundation and also evaluate the soundness and error rates for current tools and techniques. This does not merely help practitioners; it is essential for their success.

Practitioners must become more aware of the advancements of researchers, if not more involved. Researchers are producing valuable works. These works, however, are often going ignored by practitioners. If practitioners want to advance their knowledge of theoretical issues and sound methodologies, then they need to follow these researchers' works. The difficulty for practitioners is that they do not see the immediate value of digital forensics research. The research typically does not directly apply to their day-today work. As such, practitioners spend their time learning how to use techniques and tools that are of immediate value.

Practitioners and researchers should work together more closely to ensure that both understand the problems each are facing and the solutions each are producing. Some practitioners do participate in the research community. They provide researchers with more insight into the practical day-to-day problems faced in the field. They also communicate the state-of-art of digital forensics technologies to the researchers. This cross-communication empowers researchers to focus their research on practical and topical problems. It also provides practitioners with a greater understanding of theoretical issues and the work of researchers.

Venues for dialogue between researchers and practitioners do exist. The past five years have seen a consistent increase in the number of digital forensics-related books, journals, conferences, and workshops. Some are geared exclusively to practitioners, and some are geared exclusively to researchers. Most promising are those that seek a mix of researchers and practitioners [5], [3].

6. CAUSES OF CHANGE

Increased communication alone is not sufficient for improving digital, nor is it one of the most important drivers. Communication between researchers and practitioners provides each with an opportunity to better understand issues and techniques, but communication alone does not produce much change. If researchers continue to be funded to do similar work and only be evaluated by fellow researchers, then they will not likely adapt. Likewise, practitioners will not adapt unless their methods are questioned. There are two major factors in digital forensics that produce change: legal and economic.

Legal factors cause the most change in digital forensics. Several reasons exist for performing digital forensics, but the primary is to bring or respond to legal action. That is, the aim of most digital forensics investigations is to present convincing, admissible evidence to a court of law. The legal community is the primary driver for digital forensics, because they determine the overall relevance and soundness of evidence. Practitioners and researchers report to the courts as to what data and other evidence mean, and also the methods for acquiring and analyzing them. The legal community, while currently behind the times on most of these matters, determines what is admissible and reasonable based on practitioners and researchers' opinions and findings.

The four areas for which the legal community has the greatest impact are admissibility of evidence, the speed at which data must be processed, presentation of data and analysis, and standardization of investigatory techniques. The legal community demands that forensic investigators produce their findings in a standard, accepted method. This means that a common language is used, and findings and reports be presented in a consistent manner. Additionally, lawyers put demands on the reasonableness of time to perform forensic analysis.

Economic factors also play heavily in the future of digital forensics. The primary economic factor in digital forensics is that of software companies. These companies develop the commercial off-the-shelf products upon which most law enforcement and private sector investigators depend. Simple economics states that these companies' products will continue to improve based on increased competition in the digital forensics marketplace. This factor will drive companies to apply theoretical findings – such as data mining and network data analysis – in order to produce the best products. In other words, the gap between theoreticians and practitioners will be made smaller by economic factors through software manufacturers.

Economic considerations also drive researchers to produce work that is more closely aligned to what practitioners need. DARPA has had a steady decrease in funding for computer science and related fields over the past five years [7], and other funding agencies have not had an increase in funding to offset the decrease. This decrease means that researchers must compete with better research that produces practical or theoretically novel and important works. For digital forensics researchers, this means that they must produce work that is useful or otherwise important for digital forensics, instead of producing work that is never implemented, used, or referenced. Over time, this competition will force researchers to become more aware of practical and pragmatic concerns, and likewise close the gap between theory and practice.

7. CONCLUSION

Digital forensics is a young field that is being defined by its reactive nature. From its inception, digital forensics has rapidly evolved without a dearth of theoretical foundations. Practitioners have defined the best practices and developed tools on an as-needed basis, and those best practices and tools have begun to be analyzed by researchers during the past decade. This rapid development has led to many questions about the quality and soundness of those best practices and tools.

The ultimate goal of digital forensics researchers and practitioners is for the field to truly become a science like traditional forensics. Traditional forensics has an established body of literature based on

peer-reviewed, tested methodologies and techniques, which has enabled it to become a science. Digital forensics is subject to the same legal principles as traditional forensics, whereby it must satisfy the Daubert Test for admissibility. Digital forensics researchers and practitioners must thus have the same level of rigor and scientific soundness in order for digital evidence to reliably be admissible in court.

The rapid development of the field has led to a disconnect between practitioners and researchers. Practitioners constantly face new problems that they react to on an ad hoc basis. Researchers, on the other hand, face issues with less time sensitivity and seek to formally solve their problems. This scenario is similar to other scientific fields, but unlike other scientific fields, the goals for researchers and practitioners are different. Moreover, researchers and practitioners appear to have different approaches to digital forensics.

This paper lists several drivers that will solve the disconnect. The first is that both researchers and practitioners will better understand each other's goals and problems. This would allow for more relevant research and better implementation of researchers' findings. The understanding directly ties to the lack of communication. Practitioners and researchers tend to be isolated from one other, with their lack of communication. Many practitioners are unaware of the happenings in the digital forensics community, and likewise, many researchers are not aware of the capabilities of many commercial digital forensics tools. With more communication and better understanding, the digital forensics community can better and more quickly develop into a scientific field.

The other drivers are economic and legal, which will have more of an effect than communication alone. The legal community is why digital forensics was created in the first place; digital crimes or crimes involving digital evidence required a legally sound field. It is the legal community who will be the ultimate judge as to what digital forensics should achieve. Another major driver is economics. Bad research will not be funded, and bad forensics products will not be purchased. People naturally go where the money is, and as such, people in digital forensics will strive to produce the best products and research in order to get that money.

8. REFERENCES

[1] Carrier, B. 2002, "Open Source Digital Forensics Tools: The Legal Argument," @stake Reports.

[2] Daubert v. Merrell Dow Pharmaceuticals, 1993, 509 U.S. 579, pp. 92-102.

[3] Digital Forensics Research Workshop, http://www.dfrws.org, December 23, 2006.

[4] Digital Forensics Research Workshop Attendees. 2001, "A Road Map for Digital Forensic Research," Ed. Gary Palmer, Digital Forensics Research Workshop.

[5] Digital Investigations. 2005, http://www.compseconline.com/digitalinvestigation/welcome.htm, November 3, 2005.

[6] Federal Bureau of Investigations. 2005, "History of the FBI Rise of International Crime: 1980's," http://www.fbi.gov/libref/historic/history/rise.htm, December 30, 2006.

[7] Kling, J. June 2005, "DARPA and the Decline of U.S. Computer Science Research," Science Next Wave. http://nextwave.sciencemag.org/cgi/content/full/2005/06/23/1, December 5, 2006.

[8] New Technologies, Inc. January 2004, "Computer Forensics Defined," http://www.forensicsintl.com/def4.html, December 27, 2006.

[9] Palmer, G. L. 2002, "Forensic analysis in the digital world," International Journal of Digital Evidence, 1(1), http://www.ijde.org/archives/gary_article.html, November 5, 2005.

[10] Whitcomb, C.M. 2002, "An Historical Perspective of Digital Evidence: A Forensic Scientist's Point of View," International Journal of Digital Evidence, 1(1).

ABOUT THE AUTHOR

Joe Sremack is a Managing Consultant with LECG's Electronic Discovery Practice. Prior to joining LECG, Mr. Sremack conducted work and research in digital forensics, electronic discovery, and information security. He earned an M.S. in computer science from North Carolina State University and holds numerous industry certifications.